7Г

УДК 004.7:37.041

И.С. Грунский, А.С. Вороной
Государственный университет информатики и искусственного интеллекта
г. Донецк, Украина
postmaster@iai.donetsk.ua

Агентно-ориентированная система извлечения знаний из распределенных источников информации

В работе рассмотрена концепция построения многоагентной системы поиска знаний в распределенных неструктурированных информационных источниках. Предложен состав, определены функции и взаимодействия коллективов агентов при решении задачи пополнения знаний интеллектуальных многопредметных обучающих систем.

Введение

Современное развитие информационных технологий сопровождается созданием и наполнением разрозненных информационных источников и хранилищ, которые характеризуются различными способами представления информации. Автоматизация процесса извлечения знаний из больших объемов информации является одним из приоритетных направлений развития искусственного интеллекта и активно применяется при разработке систем, основанных на знаниях. Системы, основанные на знаниях, активно применяются в научных исследованиях, бизнесе, обучении. Возникновение сетевых технологий и Интернета способствует использованию этих ресурсов в качестве источников получения знаний. Однако, кроме положительных сторон, процесс получения знаний сопровождается рядом проблем, связанных с увеличением объема избыточной информации вследствие многократного дублирования, слабой структурированностью информации, представлением ее на естественном языке. Призванная решить эти проблемы концепция семантического Web в настоящее время находится на стадии развития и основная масса информационных ресурсов, накопленных в Web первого поколения, продолжает оставаться в виде, не приспособленном для поиска знаний. В связи с этим к процессу извлечения знаний предъявляется ряд требований: высокая скорость обработки больших объемов данных; гибкость относительно постоянно меняющего окружения; масштабируемость; адаптируемость к типам информационных ресурсов и их содержимому; работа с неструктурированными естественноязыковыми источниками.

Для решения проблем извлечения знаний из распределенных источников информации в настоящее время применяются технологии многоагентных систем (МАС) [1], [2] вследствие ряда преимуществ по сравнению с другими способами организации распределенных вычислений: уменьшение нагрузки на сеть; автономное и асинхронное выполнение вычислений; адаптация к условиям выполнения и т.д.

Цель работы. В данной работе рассматривается решение задачи извлечения знаний из неструктурированных источников для интеллектуальных многопредметных обучающих систем на основе экспертных систем обучения и тестирования путем использования технологии многоагентных систем.

Постановка задачи и обзор существующих подходов к ее решению

Необходимо разработать систему автоматизированного пополнения БЗ по предметным областям интеллектуальной многопредметной обучающей системы (ИМПОС) на основе информации из распределенных неструктурированных источников.

Под многоагентной системой понимается множество распределенных по сети интеллектуальных агентов, передвигающихся в поисках релевантных данных и знаний, а также кооперирующихся в процессе выработки решений [3]. Под агентом понимается система, содержащая как программные, так и аппаратные компоненты, направленная на достижение определенной цели и способная к взаимодействию со средой и другими агентами. К основным свойствам агентов относятся: способность к обучению, автономность, возможность взаимодействовать с другими агентами различными способами, способность к рассуждению (наличие механизма вывода), коммуникативность, мобильность.

При разработке архитектуры многоагентной системы различают архитектуру, поддерживающую методы взаимодействия агентов в процессе функционирования системы в целом, и архитектуру отдельного агента [2]. Архитектура взаимодействия системы агентов обеспечивает скоординированное поведение агентов: одноуровневая архитектура взаимодействия агентов (решение общей задачи в распределенном варианте) и иерархическая архитектура взаимодействия агентов (координация функционирования агентов осуществляется специальным агентом).

В состав программного агента входят два основных компонента:

1) инвариантная платформа агента (generic agent) – является общей для всех агентов и реализует функции, инвариантные к предметной области;

2) прикладная компонента – включает сценарии поведения и предметные знания определенного агента, а также спецификацию протоколов работы с ним.

Взаимодействие агентов осуществляется на основе протоколов. Протоколы описывают сценарии выполнения отдельных функций различными агентами в распределенном алгоритме решения задач. Протоколы описываются в терминах ролей, назначенных агентам. Получение агентом сообщения от одного или сообщений от нескольких агентов инициирует выполнение им некоторой функции. Сценарий обмена сообщениями задает общую модель выполнения распределенного алгоритма, исполняемого коллективом агентов.

Организация системы обновления знаний интеллектуальной многопредметной обучающей системы

ИМПОС, разрабатываемая в Государственном университете информатики и искусственного интеллекта, строится на основе принципов разработки виртуальных обучающих сред, идей реализации гиперэкспертных систем и мультиагентных технологий [4]. Предлагаемая концепция организации средств автоматизации создания и пополнения баз знаний предметных областей данной системы, основанная на использовании коллективов агентов, представлена на рис. 1. В агентно-ориентированной интеллектуальной обучающей среде (АО ИОС) процессы приобретения и обновления знаний системы обеспечиваются двумя коллективами агентов – многоагентными системами: поиска информации, содержащей знания, в распределенных удаленных источниках – МАС ПЗ и извлечения и редактирования знаний из коллекций документов по предметным областям – МАС ИРЗ.

На рис. 1 использованы следующие обозначения: ИРИ – информационные ресурсы Интернета, КД ПО – коллекции документов предметных областей, БЗ ПО – базы знаний предметных областей.

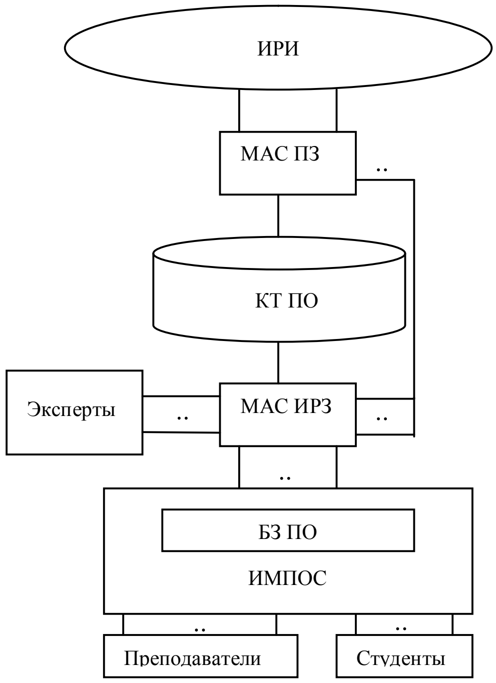


Рисунок 1 – Взаимодействие коллективов агентов поиска и извлечения знаний в АО ИОС

Использования коллективов агентов поиска и извлечения знаний особенно актуально для дисциплин, связанных с активно развивающимися предметными областями.

ИМПОС так же построена на основе агентного подхода и содержит БЗ предметных областей, БЗ обучаемых, агентов преподавателей и агентов студентов.

В состав интеллектуальных агентов МАС ПЗ входят: агент-менеджер поисковых агентов (1), агенты поисковых систем (2), агенты проверки существования документов (3), агент ранжирования документов (4), агенты получения документов (5).

Граф взаимодействия агентов МАС ПЗ при обработке запросов поиска документов для пополнения знаний в предметной области с использованием трех поисковых систем приведен на рис. 2. Вершина с номером 6 отображает агента-менеджера коллекции документов предметных областей.

Агент 1 МАС ПЗ получает запросы на поиск документов по тематике предметных областей и выдает задания агентам поисковых систем в форме запросов на поиск. В состав МАС ПЗ включены агенты поиска, использующие: поисковые системы тематических коллекций документов по предметным областям, многоцелевые поисковые системы и базы данных «скрытого» Web. Агенты проверки существования документов устанавливают актуальность ссылок, выданных поисковыми агентами. Агенты ранжирования документов исключают повторяющиеся ссылки и упорядочивают по степени релевантности ссылки на все документы, найденные используемыми поисковыми системами. Степень релевантности вычисляется на основе аннотаций, выдаваемых поисковыми системами. Менеджер поисковых агентов на основе ранжированного списка ссылок формирует задания агентам получения документов. Агенты получения документов взаимодействуют с агентом-менеджером КД ПО.

МАС ИРЗ включает агентов преподавателей – экспертов предметных областей, агентов работы с локальными знаниями, которые реализуют процедуры наполнения и корректировки баз знаний ИМПОС в процессе диалога с агентами экспертов, агентов работы с удаленными знаниями, реализующими алгоритмы автоматического извлечения знаний из коллекций документов по предметным областям.

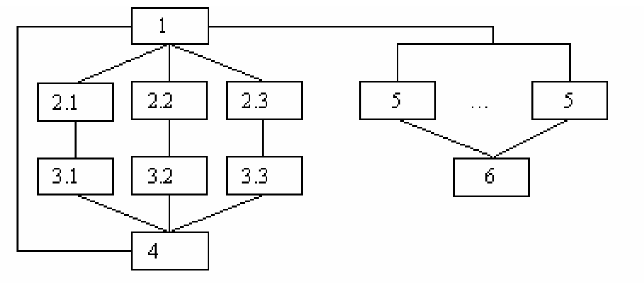


Рисунок 2 – Граф взаимодействия агентов МАС ПЗ при поиске

В режиме обработки локальных знаний пользователю – эксперту представляется удобный графический интерфейс ведения семантических сетей на основе модели «сущность-связь», средства конструирования фреймов-прототипов, создания экземпляров фреймов и другие средства представления знаний и работы с ними.

В режиме обработки удаленных знаний может производиться обновление или корректировка уже существующей базы знаний. При этом эксперт может сформировать запрос на поиск новых знаний в удаленных информационных ресурсах, анализировать знания, извлеченные агентами из КД ПО по ранее подготовленным запросам, и выполнять обновление или корректировку базы знаний.

Заключение

Предложенные типы агентов и многоагентная организация средств автоматизации создания и пополнения баз знаний экспертных обучающих и тестирующих систем реализуется в проекте интеллектуальной обучающей среды для цикла дисциплин направления подготовки студентов «Компьютерные науки».

Литература

1. Принципы построения систем для быстрой интеграции знаний из распределенных источников / [Левашова Т.В., Пашкин М.П., Смирнов А.В., Шилов Н.Г.] // Труды Международного конгресса «Искусственный интеллект в XXI веке». – Дивноморское (Россия), 2001. – Т. 1. – С. 105-119.
2. Ланин В.В. Интеллектуальное управление документами как основа технологии создания адаптируемых информационных систем / В.В. Ланин // Труды Международной научно-технической конференции «Интеллектуальные системы» (AIS'07). – М. : Физматлит, 2007. – Т. 2. – С. 334-339.
3. Agent projects. Europe's Network of Exellence for Agent-based Computing [Электронный ресурс]. – Режим доступу : http://www.agentlink.org/resources/-agentprojects-db.html, 2001.
4. Gorodetski V. Multi-agent Data Fusion Systems : Design and Implementation Issues / Gorodetski V., Karsayev O., Samoilov V. // Proceedings of the 10th International Conference on Telecommunication Systems – Modeling and Analysis, (Monterey (CA). October 3 – 6), 2002. – Vol. 2. – P. 762-774.

І.С. Грунський, О.С. Вороной

Агентно-орієнтована система добування знань із розподілених джерел інформації

У роботі розглянута концепція побудови багатоагентної системи пошуку знань у розподілених неструктурованих інформаційних джерелах. Запропонований склад, визначені функції та взаємодія агентів при розв'язанні задачі поповнення знань інтелектуальних багатопредметних навчальних систем.

I.S. Grunskiy, A.S. Voronoy

Agent-oriented system for data mining from distributed sources of information

In this work a concept of building a multi-agent system for data mining from distributed unstructured information is described. Contents, functions and interaction of groups of agents while solving a task of increasing knowledge base in intellectual learning systems is proposed.

Статья поступила в редакцию 06.08.2009.